

## 1.1 Methodology and Standards Program Area Products, Services and Customers

Computer-Assisted Survey Information Collection (CASIC) products include automated tools that reduce survey cycle time, improve data quality, and reduce survey costs, as well as an architecture that integrates a variety of tools in a seamless system. U.S. Census Bureau program areas use these products for their surveys. Other data uses also benefit from CASIC. In addition, the Computer-Assisted Survey Research Office (CASRO) staff performs initial research to find new technologies that will benefit users. CASRO creates prototypes and completes pilot projects for our customers; then, before the technology goes into production, we write transition plans to develop milestones and address subjects such as risks, roles and responsibilities.

The Integrated Statistical Laboratory's product is research used throughout the U.S. Census Bureau. Our clients are the Statistical Research Division, including the Center for Survey Methods Research; the Planning, Research and Evaluation Division, which includes the Administrative Records Research Staff; and the Computer-Assisted Survey

Research Office. Our computer resources are available to all U.S. Census Bureau Directorates, and we strongly encourage joint research projects across program areas. These projects require support and cooperation from the IT Directorate's Telecommunications Office to help program areas access and transfer files over the network. Similarly, the Integrated Statistical Laboratory supports other decentralized network staffs within the U.S. Census Bureau.

The Administrative Records Research Staff produces statistical applications of administrative records such as censuses, surveys and estimates. We supply the U.S. Census Bureau with administrative records data given to us by federal, state and local government agencies. We also produce value-added administrative records data that has been edited, geocoded, and /or modeled according to customer requirements. Our customers include the Decennial, Demographic, and Economic program areas, as well as other areas within the Methodology and Standards Directorate.

## 1.2 Methodology and Standards Program Area IT Objectives

The Methodology and Standards program area seeks to make maximum and efficient use of its IT resources to meet programmatic objectives by:

- supporting work processes that have been simplified or otherwise redesigned to reduce costs, improve effectiveness, and make maximum use of commercial, off-the-shelf technology;
- producing timely, relevant, and quality products to meet customer's needs;
- develop innovative solutions to methodological and statistical problems to maintain and enhance the U.S. Census Bureau's position as the premiere statistical agency;
- develop approaches to minimize respondent burden and encourage respondent participation;
- develop a diverse workforce that is widely recognized for its excellence in survey and statistical research, methodology, computational expertise, and technological innovation;
- augment existing corporate standards to ensure quality data products;
- attain quality products through re-engineering processes; and
- develop prototype system for a unified metadata, accessed electronically.

## 2.1 IT Systems Description

### 2.1.1 Detailed Description of Computer-Assisted Survey Information Collection (CASIC)

The U.S. Census Bureau established the Computer-Assisted Survey Research Office (CASRO) to investigate and evaluate technology advances in electronic information collection and provide the stimulus and infrastructure to support U.S. Census Bureau programs. CASRO's scope includes the integrated data collection, capture, cleaning, and post-collection processing for mandated U.S. Census Bureau surveys and censuses as well as reimbursable surveys conducted for external customers. These statistics are used to produce government indexes, leading indicators that track how well our nation's economy is doing; some examples are:

- employment statistics and producer price indexes for the Bureau of Labor Statistics;
- gross domestic products for the Bureau of Economic Analysis; and
- industrial production and capacity use for the Federal Reserve Board.

Using CASIC tools to integrate data collection, capture, cleaning and post-collection processing potentially affects all U.S. Census Bureau business processes. CASRO's staff (with support from the IT Directorate) is working with other program areas to examine and possibly re-engineer these business processes to use an integrated toolbox approach. CASRO's approach calls for re-engineering data collection, post-collection processing, and dissemination.

#### Data Collection

Data collection consists of the following processes:

- **Project Initiation and Management;**
- **Universe Creation and Sample Selection;**
- **Preparation for Data Collection; and**
- **Data Collection and Entry.**

CASRO's staff is working with the program areas to develop requirements and evaluate vendor software to accommodate some of these functions.

**Project Initiation and Management:** this is the process for formulating data collection methodology, cost estimation, clearance process, procedure and materials development as well as monitoring of costs, quality, and progress.

Historically, the resources needed to initiate and manage a U.S. Census Bureau data collection effort have varied widely and were generally approached on an ad-hoc, custom basis. A variety of commercial off-the-shelf management information systems software was normally used, but no standard progress reporting system existed. Payroll information from field representatives was collected through the mail. Managing case assignments and transmissions was also customized, with varying degrees of automation and controls. Standards were rarely used and, while a phased lifecycle approach to systems design was encouraged, the commitment of resources and standard tools to accomplish this did not exist.

effort to modernize our survey mail preparation activities by re-engineering methodologies and applying advanced technologies. STAMP's goal was to establish a high volume, electronic printing facility operating within a centrally controlled, state-of-the-art mail processing environment. Together, these printing and processing components form a technologically advanced, fully integrated print-on-demand mailing system that will meet U.S. Census Bureau printing, publishing, and mail processing needs well into the next century.

Designing and generating computer-assisted questionnaires (instruments) involves developing specifications with the sponsoring office/agency, authoring the survey instrument, and transmitting the instrument to the field representatives or data capture staff.

The CASIC plan includes the continued move to computer-assisted questionnaires and standardizing the steps involved in setting up and authoring an automated collection instrument. The Computer-Assisted Survey Research Office will include tools to simplify setting up instruments through user interfaces.

Further automating survey instruments will reduce the risk of printing errors, reduce the turnaround time in correcting or fine-tuning the wording of the questions, and facilitate using previously reported information in the collection process itself. Paper and pencil surveys that would benefit from previous report/collection information often resort to re-asking for the information or preprinting the information (which is costly) on particular forms.

The Master Control System is a centralized control and tracking system for CASIC's data collection components. Currently, the Master Control System provides case

management functions for Computer-Assisted Telephone Interviewing (CATI), Computer-Assisted Personal Interviewing (CAPI), Computer-Assisted Data Entry, and outbound faxing. It allows cases to be transferred between CATI and CAPI. In the future, we will expand the Master Control System to include other CASIC data collection technologies and provide survey status information to survey managers.

**Data Collection and Entry:** this involves some type of respondent interaction that results in raw information; receipts processing, which may include check-in, screening, preliminary editing, data capture and coding; and resolving problems.

U.S. Census data collection systems combine automated and manual processes that we are enhancing by expanding CASIC. To date, we have concentrated the development of CASIC tools on data collection and capture. We can eliminate data keying from any statistical program that uses computer-assisted interviewing (CAI) technology, since CAI combines collection and entry into one step. Using CAI can also eliminate some of the preparations associated with paper forms, as well as many of the subsequent editing and corrections.

Many of the U.S. Census Bureau's statistical programs continue to use paper and pencil for data collection and keying for data entry. Collected forms generate a data entry workload of over 37 million records in a typical year at the National Processing Center and the Regional Offices.

While automating questionnaires represents a CASIC goal and effectively combines data entry with data collection, the U.S. Census Bureau's efforts in these areas since the early 1980s were uncoordinated and approached on an ad-hoc basis by survey or program.

The U.S. Census Bureau has expanded its authoring staff to accommodate the ever-increasing number of surveys to be automated. Another undertaking of this group is to standardize certain parts of instruments so that common variables and question screens can be used for introducing surveys, collecting household information, etc. Long-range plans include developing an authoring interface that would simplify the setup, coding, and documentation of automated instruments regardless of the data collection methodology they use.

**Computer-Assisted Personal Interviewing (CAPI):** in January 1994, the Current Population Survey was the first major survey to use CAPI. Since then, many new surveys have implemented CAPI; several major, older surveys have converted or are in the process of converting from paper and pencil data collection systems to CAPI. Additionally, the Census 2000 Accuracy and Coverage Evaluation program will use CAPI. For more specific information, please see the Decennial and Demographic program areas' sections of this Plan.

We are also keeping abreast of pen-based technologies for laptops. If the U.S. Census Bureau determines these systems feasible for production use, they would represent a new generation of CAPI capabilities, integrating data collection modules with area maps and locating procedures.

**Computer-Assisted Data Entry:** the CASIC plan recognizes that not all data collection efforts can be fully automated and combine data collection and capture into one process. There will still be a need for paper and pencil forms with attendant data entry systems.

Currently the National Processing Center (NPC) in Jeffersonville, Indiana performs data keying. In an average year, NPC keys over 37 million records, with these numbers rising by nearly 100 million when the Economic Census is over. Currently, NPC uses two data entry systems: DIGITAL VAX (which the U.S. Census Bureau used for the 1997 Economic Census), and networked PCs using KeyEntry III software (this replaced the TARTAN data entry system). The KeyEntry III system has the capability for both "heads down" and "heads up" keying. "Heads down" keying means that the data is keyed quickly with no concurrent editing beyond basic edits such as range checks; any coding is usually done prior to keying. "Heads up" data entry means that edits and/or coding are done at the time of keying.

The CASIC environment includes a multi-functional data capture facility at the National Processing Center. The facility consists of 250 networked microcomputer workstations connected to a centralized control system. The facility manages heads-up and heads-down data entry, Computer-Assisted Telephone Interviewing, and Touchtone Data Entry/Voice Recognition Entry. We will eventually use this facility as a receiving point for data reported by respondents using Computerized Self-Administered Questionnaires and other technologies.

**Touchtone Data Entry/Voice Recognition Entry:** this technology is in production and ready to accept additional work. Small vocabulary Voice Recognition Entry is incorporated into the Touchtone Data Entry system; we are still researching medium and large vocabulary capabilities.

**Imaging and Workflow:** CASRO is developing an imaging system that the U.S. Census Bureau will use to mail and receive current surveys. CASRO is also researching workflow software for data capture and other survey processes. Our plan is to build upon the Decennial program area imaging application to meet the requirements for Demographic and Economic surveys and Economic censuses.

CASRO has developed and installed a pilot imaging, automated data capture, and workflow system. The initial program on this system is the 1998 Annual Survey of Manufactures. This system uses a Kodak 9500 scanner (as does the Decennial program area's Data Capture System 2000) to scan incoming questionnaires and correspondence. The system stores the questionnaires and correspondence as images, using Feith Document Database software from Feith Systems and Software. These images are immediately available for retrieval to analysts' desktops at the National Processing Center and at Headquarters. Microsystems Technology's Optical Character Recognition for Forms software extracts data from the questionnaires; any questionable characters are verified/corrected by a key-from-image clerk.

CASRO will undertake a detailed evaluation of this system to determine how effective the automated data capture (optical recognition with key-from-image verification) is and what are the benefits of using imaging technology to store images for future retrieval. Future studies include the inbound fax and workflow capabilities of the system. Programs expressing interest in using this system include the American Community Survey, the Jeffersonville Activity Reporting System, and the Manufacturing, Shipments Orders, and Inventory (M3) survey.

**Paperless Fax Imaging Reporting System:** the Paperless Fax Image Reporting System has both outbound (faxing replacement forms to respondents) and inbound (receiving completed forms from respondents via fax) component. The Technologies Management Office is already maintaining a production system for outbound faxing and CASRO has written a transition plan for this system. CASRO has also written a research plan for an inbound system and plans to write a transition plan. A system is currently being tested for M3; this system uses commercial off-the-shelf software packages for its three main software components: the fax receiving and routing software, the software used for the Correct From Image operation, and the image repository setup. It receives faxes from a sample of M3 respondents through a toll-free number, reads them using intelligent character recognition technology, routes the images to a program, and then sends the data through a load file to the production M3 database. A centralized system would allow the U.S. Census Bureau to maintain fewer toll-free numbers and reduce some duplication of staff responsibilities at Headquarters and the National Processing Center. It would absolve the individual branches from maintaining their own systems while still delivering data and check-in information to Headquarters quickly as surveys near their closeout dates. In the future, we will integrate the inbound Paperless Fax Imaging Reporting System with the imaging system currently being developed.

(IIS) program is an effort to reorganize the U.S. Census Bureau's data creation and delivery systems so that we can deliver "integrated information solutions" to our customers inside and outside the Bureau. IIS is an answer to a customer's question, delivered without the customer's first having to learn how the U.S. Census Bureau, our programs, or our data files are organized. The DADS project was a starting point for the IIS program. In June 1998, a cross-divisional/cross-functional team was established to consider the future of data access and dissemination at the U.S. Census Bureau. The team is using the business process reengineering methodology developed by Canal Bridge Consulting and supported by the CASRO staff. The work of the full team was completed in January 1999; work yet to be completed includes beginning detailed implementation planning with pilot demonstration projects.

The IIS program will eventually incorporate the data delivery component of DADS, along with FERRET and other dissemination vehicles. The IIS program will develop new software and procedural tools to allow U.S. Census Bureau staff to more efficiently develop and disseminate data and help customers. While we hope to get a head start with pilot or demonstration projects (Phase 1 is from February 1999 through September 2000), much of our work and development will have to wait until we have funding (Phase 2, starting October 2000). We will put many of those systems in place within a year or two after that, even as we continue development to meet new challenges (Phase 3, starting in 2002 or 2003).

Our Census Modernization initiative supports all the activities discussed above. Those activities, in turn, support program areas across the U.S. Census Bureau.



**CASIC Milestones, FY 01**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Workflow Test	01/00	12/00			
Imaging Transition Plan	01/00	12/00			
CSAQ Internet Transition Plan	06/00	06/01			
Computer-Assisted Telephone/Personal Interviewing Graphic User Interface Research	03/98	09/01			

**CASIC Milestones, FY 02**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Migration of CATI/CAPI Surveys to Windows (Phase 1)	09/01	09/02			
Migration of Surveys to Imaging (Phase 1)	01/00	09/02			
Migration of Surveys to CSAQ Internet (Phase 1)	01/00	09/02			

**CASIC Milestones, FY 03**

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Development of Integrated Computing Environment Tools	10/00	09/03			

#### 2.1.4. Computer-Assisted Survey Information Collection (CASIC) Risks

Before moving any new technology to a production environment, we prepare a transition plan that includes the technology's purpose, cost/benefit component, production environment, resource needs, training needs, funding, divisional responsibilities and schedule. These transition plans discuss the

risks associated with implementing the specific technology. Various customers assume the risks, and the production areas' Information Technology Plans discuss the risks further. Please see section 2.1.5, below, for supporting documents.

#### 2.1.5 Computer-Assisted Survey Information Collection (CASIC) References

CASIC is supported by the following planning documents:

- Budget submission for FY 2001, dated June 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 83-84, 99;
- Integrated Computing Environment White Paper, dated September 19, 1997;
- *Computer-Assisted Survey Information Collection*, Requirement Initiative DR01-9401, approved August 25, 1994, renewed July 6, 1998 through 2002;
- Transition Plans for STAMP, CSAQ, PFIRS, and Metadata; and
- the following Business Plans:
  - Technology: Computerized Self-Administered Questionnaires;
  - Technology: Develop Transition Plan for Paperless FAX Image Reporting System;
  - Technology: Establish a Re-engineering Services Component to CASRO; and
  - Re-engineering: Conduct Process Improvements for Three Prototype Projects.

#### 2.2.1. Detailed Description of Integrated Statistical Laboratory

The Integrated Statistical Laboratory provides a statistical research computing environment with the power and flexibility to handle the following:

- multiple operating systems;
- many users;
- large files and databases;
- a large variety of statistical, analytical, and graphical packages and libraries;
- highly computer-intensive customized programs; and
- the network communications to permit researchers to share information easily and accurately.

Our Sun cluster and PC network systems have evolved continuously over the last ten years with yearly technology upgrades and phased retirement of older systems. These systems are part of the basic computing infrastructure that supports research in the Methodology and Standards Directorate.

The Integrated Statistical Laboratory conducts broad research and development activities in the following areas:



that vary substantially in performance, disk capacity, and display capabilities. These workstations provide CPU and local storage dedicated to individual projects, as well as a high-resolution graphical display.

We also have 20 X-terminals that are a substantially less expensive alternative to workstations, yet provide a highly responsive X-window environment that communicates easily with any VAX or UNIX machine on the network. We use XonNet software on some PC desktops to also provide access to X-windows.

The Integrated Statistical Laboratory UNIX platforms provide programming and soft-

ware development tools, data analysis software, graphical packages for statistics and presentations, and advanced mathematical and statistical software. These software tools are critical parts of nearly every research project, and it is essential for effective research that these be highly capable.

The HP server is an HP730 processor dedicated to supporting the Industry and Occupation coding project, which is shared by the Housing and Household Economic Statistics Division and the Statistical Research Division.

### 2.2.2 Integrated Statistical Laboratory Progress Against Planned Milestones

Integrated Statistical Laboratory Milestones, FY 98					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/97	09/98			Completed.

**Progress to date:** Consistent with its open systems philosophy, the Integrated Statistical Laboratory has continued to upgrade its hardware and software resources. The Integrated Statistical Laboratory continues to provide maintenance contracts for both hardware and software.

Integrated Statistical Laboratory Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/98	09/99			In progress.
Integrated Statistical Laboratory Upgrade Storage	10/98	09/99			In progress.

### 2.2.3 Integrated Statistical Laboratory Performance Measure

Integrated Statistical Laboratory Performance Measure			
Performance Goals	Performance Measures	Target Performance	Current Performance
Use the systems to start research	% of program areas and others who adapt statistical methods and software	100%	100%

### 2.2.4 Integrated Statistical Laboratory Risks

There are significant risks to the continued success of the Integrated Statistical Laboratory; these risks are, primarily, loss of skilled systems staff and loss of funding. Another risk is the loss of the Statistical Research Division computer room in Federal Office Building 4. This room holds the crit-

ical Statistical Research Division UNIX servers. No real adequate backup systems exist for these servers, although we could use other U.S. Census Bureau UNIX servers to run critical applications. We will protect backup data by placing it in a safe in the secure area at the Bowie Computer Center.

### 2.2.5. Integrated Statistical Laboratory References

The Integrated Statistical Laboratory is supported by the following planning documents:

- Budget Submission for FY 2001, dated June 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 82-84, 99;
- *Integrated Statistical Laboratory*, Requirement Initiative SM02-9801, approved April 28, 1998; and
- the Annual Report of the Statistical Research Division.

### 2.3.1 Detailed Description of Administrative Records Research System

This research program's goal is to expand statistical uses of administrative records to reduce respondent burden and costs, improve coverage of censuses and surveys, and generate new information not otherwise available from surveys or censuses. Before being deemed statistically reliable, administrative records data must undergo extensive research to address data quality, consistency with census definitions, and coverage. This research supports uses of administrative

records for decennial, intercensal estimates, survey and address list applications.

Although the primary purpose of the Administrative Records Research System (ARRS) is to support research, the U.S. Census Bureau is planning several program applications—most notably for Census 2000, Master Address File Updates, and the intercensal estimates programs—that will require ARRS' resources. Census 2000 will use administrative records to conduct an experi-

The core Administrative Records Research System processing system now consists of two VAX Alphaserver 8400 5/625 systems; each system has eight CPUs and four GB of main memory. One system (DMCA01) is configured to operate in VMS and has one terabyte of RAID storage. The second system (DMCA01) is configured to operate in UNIX and has two terabytes of RAID storage. Each system also has a TL893 264 cartridge/3 drive DLT tape library. The two

DLT tape libraries can hold a maximum of 18.4 terabytes (native) of data.

The primary reason we have configured these machines under two different operating systems is the difficulty in porting hundreds of FORTRAN programs from DEC VMS to DEC UNIX. We are beginning a gradual phased migration from DEC VMS to DEC UNIX.

### 2.3.2 Administrative Records Research System Progress Against Planned Milestones

Administrative Records Research System Milestones, FY 98-03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
No milestones to report					

### 2.3.3 Administrative Records Research System Performance Measures

Administrative Records Research System Performance Measures			
Performance Goals	Performance Measures	Target Performance	Current Performance
Reduce respondent burden, reduce data collection costs and more timely statistics	% research results delivered to customers and policy makers on schedule	100%	N/A

### 2.3.4 Administrative Records Research System Risks

Our research program would be severely compromised if access to key files, such as the Internal Revenue Service (IRS) Individual Master File or the Information Return Master File, were eliminated. In accordance with our memorandum of understanding with the IRS, they will be performing a site audit in 1999; this site audit includes re-

viewing our core processing systems, our internal customer's systems, and the systems in the Economic program area. If the results of this review are negative, our ability to access IRS data agency-wide could be limited or even eliminated.

The success of our research program is dependent on our ability to hire and retain

### 3.2 Methodology and Standards Program Area Infrastructure Progress Against Planned Milestones

Methodology and Standards Infrastructure Milestones, FY 98-03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
No milestones to report					

### 3.3 Methodology and Standards Program Area Infrastructure Performance Measures

The Methodology and Standards infrastructure has no performance measures to report.

### 3.4 Methodology and Standards Program Area Infrastructure Risks

The Statistical Research Division has a single Novell server for its office automation functions; a failure could cause serious problems. We will mitigate this by purchasing a new server and using the current server as a backup device.

### 3.5 Methodology and Standards Program Area Infrastructure References

The Methodology and Standards program area infrastructure is supported by the following planning documents:

- *Enterprise Printer Resources*, Requirement Initiative PRMAP9701, approved May 13, 1998; and
- *Enterprise Personal Computer Management and Acquisition Plan*, Requirement Initiative PCMAP9601, approved December 5, 1995.